PROVIDING TRANSLATIONS ENCODED WITHIN EMBEDDED DIGITAL INFORMATION

Inventor(s):

Thomas E. Creamer

Peeyush Jaiswal

Victor S. Moore

International Business Machines Corporation

IBM Docket No. BOC9-2003-0092

IBM Disclosure No. BOC8-2003-0117

PROVIDING TRANSLATIONS ENCODED WITHIN EMBEDDED DIGITAL INFORMATION

BACKGROUND

Field of the Invention

[0001] The invention relates to speech or voice translation systems.

Description of the Related Art

[0002] Spoken language is typically the most natural, most efficient, and most expressive means of communicating information, intentions, and wishes. Speakers of different languages, however, face a formidable problem in that communication is thwarted unless the language barrier is removed. As the global economy brings together persons of various nationalities, a forum is needed that provides efficient and accurate communication, which effectively eliminates the language barrier.

[0003] Translation systems have emerged to address this need. Presently available translation systems are capable of receiving a speech signal in a first language. Typically, the speech signal is provided to a speech recognition system to determine a textual transcript from the speech signal. The textual transcript then can be processed or translated into a different language, for example through the use of a translation system such as one using natural language processing. The resulting translated text then can be provided to another person or device as text or played through a text-to-speech system.

SUMMARY OF THE INVENTION

[0004] The present invention provides a method, system, and apparatus for including transcription information within a voice stream or speech signal. One aspect of the present invention can include a method of providing a translation within a voice stream. The method can include receiving a speech signal in a first language, determining text from the speech signal, and translating the text to a second and different language.

[0005] The method further can include encoding the translated text within the speech signal. For example, the encoding step can include the translated text within the speech signal as digital information. The resulting speech signal can specify both speech in the first language and a textual translation of the original speech in the second and different language. The encoding step can include removing inaudible portions of the voice signal and embedding the translated text in place of the inaudible portions of the speech signal.

[0006] Another embodiment of the present invention can include transmitting the resulting speech signal. The speech signal specifying the translated text can be received and the translated text can be decoded. Accordingly, a representation of the translated text can be presented. Additionally, an audible representation of the received speech signal can be played. Notably, the audible representation of the received speech signal can be played substantially concurrently with the presentation of the translated text.

[0007] Other embodiments of the present invention can include a system having means for performing the various steps disclosed herein and a machine readable storage for causing a machine to perform the steps described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] There are shown in the drawings, embodiments which are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown.

[0009] FIG. 1 is a schematic diagram illustrating a system for providing a translation within an audio stream in accordance with the inventive arrangements disclosed herein.

[0010] FIG. 2 is a flow chart illustrating a method of providing a translation within an audio stream in accordance with the inventive arrangements disclosed herein.

DETAILED DESCRIPTION OF THE INVENTION

[0011] FIG. 1 is a schematic diagram illustrating a system 100 for providing a translation within a voice stream in accordance with the inventive arrangements disclosed herein. As shown, the system 100 can include a speech recognition system 110, a translation system 120, and an encoder 130.

[0012] The speech recognition system 110 can receive digitized speech signals 105 and produce a textual representation from the speech signals. That is, the speech recognition system 110 can convert received speech to text 115. Notably, the speech recognition system 110 can time stamp the recognized text 115 so that the text 115, or a derivative thereof, can be aligned with the original speech signal 105 at a later time. The speech recognition system 110 can provide the original speech signals 105 to the encoder 130. The speech recognition system 110 also can time stamp the speech signals 105 provided to the encoder 130.

[0013] The translation system 120 can translate the text 115 to a second and different language to produce a translation 125, which is a textual translation of text 115. The translation system 120 also can preserve any timing information that may be included within the recognized text 115 provided by the speech recognition system 110.

[0014] The encoder 130 can receive both the speech signals 105 and the translation 125. The encoder 130 can encode the text of the translation 125 into the speech signal 105, resulting in speech signal 135 having embedded digital information specifying a textual representation of the speech signal 105, where the textual representation is in a different language than the original speech.

[0015] More particularly, one aspect of the encoder 135 can be implemented as a perceptual audio processor, similar to a perceptual codec, to analyze the received speech signal 105. A perceptual codec is a mathematical description of the limitations of the human auditory system and, therefore, human auditory perception. Examples of perceptual codecs can include, but are not limited to MPEG Layer-3 codecs and MPEG Layer-4 codecs. The encoder 135 is substantially similar to the perceptual codec with the noted exception that the encoder 135 can, but need not implement, a second stage of compression as is typical with perceptual codecs.

[0016] The encoder 135, similar to a perceptual codec, can include a psychoacoustic

model to which source material, in this case the speech signal 105, can be compared. By comparing the speech signal 105 with the stored psychoacoustic model, the perceptual codec identifies portions of the speech signal 105 that are not likely, or are less likely to be perceived by a listener. These portions are referred to as being inaudible. Typically a perceptual codec removes such portions of the source material prior to encoding, as can the encoder 135. The encoder 135, however, adds the translation 125 as embedded digital information in place of the removed inaudible portions of the speech signal 105.

[0017] Still, those skilled in the art will recognize that the present invention can utilize any suitable means or techniques for digitally encoding the translation 125 and embedding such digital information within a digital voice stream or speech signal. As such, the present invention is not limited to the use of one particular encoding scheme.

[0018] FIG. 2 is a flow chart illustrating a method 200 of providing a translation within a voice stream in accordance with the inventive arrangements disclosed herein. The method can begin in step 205 where speech is received by the speech recognition system. As noted, the speech can be provided to the speech recognition system in digitized form and can be in a first language, such as English.

[0019] In step 210, the speech recognition system can convert the received speech to text. The speech recognition system further can provide the original speech signals as output to the encoder. As noted, the recognized text, as well as any speech provided from the speech recognition system can be time stamped so that recognized text, whether translated or not, can later be aligned with the original speech. In step 215, the text provided from the speech recognition system can be translated to a second and different language.

[0020] In step 220, the translated text can be encoded into the original speech. That is, the translated text can be embedded within the voice stream of the original speech. Accordingly, the original speech remains in the first language, for example English, while the encoded translated text is in a second and different language such as French or Japanese. Notably, the encoded translation can, but need not, be synchronized with the original speech when encoded.

[0021] The translation can be sent to another destination as an encoded stream of

digital information embedded within the digital voice stream or speech signal. The encoder can identify which portions of the received speech signal are inaudible, for example using a psychoacoustic model. For instance, humans tend to have sensitive hearing between approximately 2 kHz and 4 kHz. The human voice occupies the frequency range of approximately 500 Hz to 2 kHz. As such, the encoder can remove portions of a speech signal, for example those portions below approximately 500 Hz and above approximately 2 kHz, without rendering the resulting speech signal unintelligible. This leaves sufficient bandwidth, in the case of a telephony voice stream, within which the translation can be encoded and sent. Still, it should be appreciated that other frequency ranges may be more optimal depending upon the bandwidth of the transmission channel.

[0022] The encoder further can detect sounds that are effectively masked or made inaudible by other sounds. For example, the encoder can identify cases of auditory masking where portions of the speech signal are masked by other portions of the speech signal as a result of perceived loudness, and/or temporal masking where portions of the speech signal are masked due to the timing of sounds within the speech signal.

[0023] It should be appreciated that as determinations regarding which portions of a speech signal are inaudible are based upon a psychoacoustic model, some users will be able to detect a difference should those portions be removed from the speech signal. In any case, inaudible portions of the speech signal can include those portions of the speech signal as determined from the encoder that, if removed, will not render the speech unintelligible or prevent a listener from understanding the content of the speech signal. Accordingly, the various frequency ranges disclosed herein are offered as examples only and are not intended as limitations of the present invention.

[0024] The encoder can remove the identified portions, i.e. those identified as inaudible, from the speech signal and add the translation in place of the removed portions of the speech signal. That is, the encoder replaces the inaudible portions of the speech signal with digital translation information.

[0025] In step 225, the resulting speech or voice stream, having translated text embedded therein, can be sent or transmitted to another destination or device. The

resulting voice stream can be sent over any of a variety of different communications channels including, but not limited to, a telephony link, whether conventional or IP-based, a wireless communications channel, or the like.

[0026] In step 230, the other device can receive the speech and embedded translated text. The receiving device, or another device communicatively linked to the receiving device, can decode the embedded translated text in step 235. In step 240, the receiving device can present the embedded translated text. For example, the translated text can be presented visually or can be played audibly, for instance through a text-to-speech system. In step 245, the original speech in the first language can be played audibly. In one embodiment of the present invention, the presentation of the translated text and the playing of the original speech can occur substantially simultaneously. As both the translated text and the speech can include time stamp information, the presentation of both can be synchronized.

[0027] The inventive arrangements disclosed herein have been presented for purposes of illustration only. As such, the various examples presented herein should not be construed as a limitation of the present invention. For example, the particular languages used are not intended as a limitation on the present invention as the speech recognition and translation systems can operate on any of a variety of different languages. Further, in another embodiment, the present invention can provide an embedded transcript within the speech that is in the same language as the speech signal. In that case, rather than providing the text determined from the speech recognition system to the translation system, the text can be provided directly to the encoder to be embedded within the original speech signal or voice stream.

[0028] The present invention can be realized in hardware, software, or a combination of hardware and software. The present invention can be realized in a centralized fashion in one computer system, or in a distributed fashion where different elements are spread across several interconnected computer systems. Any kind of computer system or other apparatus adapted for carrying out the methods described herein is suited. A typical combination of hardware and software can be a general purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein.

[0029] The present invention also can be embedded in a computer program product, which comprises all the features enabling the implementation of the methods described herein, and which when loaded in a computer system is able to carry out these methods. Computer program in the present context means any expression, in any language, code or notation, of a set of instructions intended to cause a system having an information processing capability to perform a particular function either directly or after either or both of the following: a) conversion to another language, code or notation; b) reproduction in a different material form.

[0030] This invention can be embodied in other forms without departing from the spirit or essential attributes thereof. Accordingly, reference should be made to the following claims, rather than to the foregoing specification, as indicating the scope of the invention.